Seasonal Variation in Surgical Outcomes as Measured by the American College of Surgeons-National Surgical Quality Improvement Program (ACS-NSQIP)

Michael J. Englesbe, MD, Shawn J. Pelletier, MD,* John C. Magee, MD,* Paul Gauger, MD,* Tracy Schifftner, MS,† William G. Henderson, PhD,† Shukri F. Khuri, MD,‡ and Darrell A. Campbell, MD**

Objective: We hypothesize that the systems of care within academic medical centers are sufficiently disrupted with the beginning of a new academic year to affect patient outcomes.

Methods: This observational multiinstitutional cohort study was conducted by analysis of the National Surgical Quality Improvement Program–Patient Safety in Surgery Study database. The 30-day morbidity and mortality rates were compared between 2 periods of care: (early group: July 1 to August 30) and late group (April 15 to June 15). Patient baseline characteristics were first compared between the early and late periods. A prediction model was then constructed, via stepwise logistic regression model with a significance level for entry and a significance level for selection of 0.05.

Results: There was 18% higher risk of postoperative morbidity in the early (n = 9941) versus the late group (n = 10313) (OR 1.18, 95%, CI 1.07–1.29, $P = 0.0005$, c-index 0.794). There was a 41% higher risk for mortality in the early group compared with the late group (OR 1.41, CI 1.11-1.80, $P = 0.005$, c-index 0.938). No significant trends in patient risk over time were noted.

Conclusion: Our data suggests higher rates of postsurgical morbidity and mortality related to the time of the year. Further study is needed to fully describe the etiologies of the seasonal variation in outcomes.

(*Ann Surg* 2007;246: 456–465)

A surgical encounter is a complex undertaking. It follows that there is a strong association between depth of operative experience, at both an institutional and individual level, and good surgical results.[1–3] Likewise, there is literature which documents the importance of team training and coordinated experience in producing favorable outcomes in aviation safety[4,5] and the emergency room environment.[6] In most academic medical centers, July and August are marked by an influx of relatively inexperienced trainees, unfamiliar with their roles and responsibilities. We hypothesized that a disruption of hospital systems in July and August (the July effect) would be of sufficient magnitude to adversely influence surgical results. If adverse clinical outcomes could be related to a seasonal disruption of hospital systems, attention to this phenomenon would be an important priority for the health care community.

A seasonal variation in surgical results, such as we have postulated, has not been previously demonstrated.[7–13] Presumably, this is because existing quality metrics in surgery have not been sufficiently standardized and risk adjusted to allow valid month-to-month comparisons on a large scale basis. A multicenter risk-adjusted analysis with sufficient statistical power to detect differences became possible with the institution of the American College of Surgeons-National Surgical Quality Improvement Program (ACS-NSQIP), the first nationally validated outcome-based and risk-adjusted program for the enhancement of surgical quality.[14–16] The ACS-NSQIP system uses a set of defined comorbidities and endpoints (30-day mortality and morbidity) and a rigorous risk-adjustment system to allow for comparison of results among hospitals. Importantly, the ACS-NSQIP system adjusts for some of the confounding variables which could influence an analysis of seasonal variation. Such factors as seasonal variations in the complexity of the surgical procedures, the degree of patient illness, and the emergency status of cases could, if not accounted for, obscure an effect due to a seasonal breakdown in hospital systems.

In this study, we analyzed the records of over 60,000 patients enrolled in the ACS-NSQIP in 14 academic medical centers and 4 large private, community-based hospitals over a 3-year period (Table 1). Our observations demonstrate a significant increase in operative mortality and morbidity in the early (July-August) versus the late (May-June) periods of the academic year.

From the *Department of Surgery, University of Michigan Health System, Ann Arbor, MI; †Department of Veterans Affairs, VA—National Surgical Quality Improvement Program, Aurora, CO; and ‡National Surgical Quality Improvement Program, West Roxbury, MA.

Dr Englesbe is supported by a grant form the American Surgical Association. Dr Campbell is supported by a grant from Blue Cross and Blue Shield of Michigan.

Reprints: Michael J. Englesbe, MD, University of Michigan Health System, 2926 Taubman Center, 1500 E Medical Center Drive, Ann Arbor, MI 48109-0331. E-mail: englesbe@med.umich.edu.

ISSN: 0003-4932/07/24603-0456
DOI: 10.1097/SLA.0b013e31814855f2

TABLE 1. The 14 Academic and 4 Community Medical Centers Participating in the ACS-NSQIP Patient Safety in Surgery Study

University of Michigan	University of Utah	Cornell University
University California—San Francisco	The Brigham and Women's Hospital	Columbia University
Massachusetts General Hospital	Washington University (St Louis)	University of Florida
St Louis University	Emory University	University of Virginia
University of Maryland	University of Kentucky	Newton—Wellesley Hospital
North Shore Medical Center	Puritan Medical Center	Faulkner Hospital

METHODS

Subjects

Permission for this study was obtained from the Institutional Review Board for Research at the University of Michigan Health System. This observational multiinstitutional cohort study was conducted by analysis of the NSQIP-PSS database. Details of the NSQIP-PSS database data definitions and policies are well described and we followed identical methodology.[16–19] Of note, data were collected starting in 2001, but 80% of the cases were collected in 2003 and 2004. The 4 community hospitals contributed a relatively small number of cases because of their late enrollment data in the ACS-NSQIP. No cases were eliminated from the analysis.

The NSQIP collects preoperative risk factors, variables about the operation the patient undergoes, and data on mortality and surgical complications up to 30 days after the operation. The patient population consists of patients undergoing major operations at the participating sites under general, spinal, or epidural anesthesia. Minor operations of known low morbidity and mortality are excluded. Also, patients undergoing operations in the past 30 days are excluded. In addition, very common operations such as inguinal hernia repairs and breast lumpectomies are limited to the first 5 in each 8 day cycle. To obtain a representative sample of operations, the first 40 consecutive eligible operations are entered into the NSQIP in each 8 day cycle, with each cycle starting on a different day of the week. The NSQIP in the private sector is currently limited to major vascular and general surgery operations. Preoperative, operative, and postoperative variables were selected from the data collection in the NSQIP protocol on the basis of clinical relevance, reliability of the data collection, and availability and ease of collection. Because the intention of the NSQIP was to cover all major operations, the design of the study called for mostly generic variables, and few disease specific or operation specific variables. A surgical nurse reviewer is assigned at each medical center to collect the NSQIP data. The nurses receive in depth training on the study protocol, patient selection, definition of the variables, and data collection methods. An operations manual is also provided to each nurse that outlines these procedures in depth. Regular conference calls and annual nurse meetings are held to maintain data uniformity and accuracy. The nurses are coordinated at the national level, and interrater reliability site visits are made at each participating site annually.

Patients were divided into 2 groups: those who had their operation from July 1 to August 30 (Early group) and those who had their operation from April 15 to June 15 (Late group). We prospectively decided not to include the final 2 weeks in June in the late group because of concerns that this period may be a transition period for hospital personnel staffing. Morbidity was treated as a dichotomous variable with patients categorized as having had a morbidity event if they had any (1 or more) of the 19 adverse events defined within the NSQIP data points.[20] The mortality period is the first 30 postoperative days.

Statistical Analysis

The preoperative characteristics and intraoperative information were categorized and compared between the 2 groups. χ^2 analyses were used for categorical data and *t* tests for continuous data to look for statistically significant differences. Multivariable regression modeling was used to identify significant independent risk factors for morbidity and mortality. Potential independent variables included clinically significant data on demographics, comorbid conditions, preoperative laboratory values, intraoperative events, and early or late time of year. A prediction model was then constructed, via stepwise logistic regression model with a significance level for entry and a significance level for selection of 0.05. A linear regression of the unadjusted morbidity and mortality rates by month for the full year was completed. Statistical analyses were performed with SAS software, version 9.1 (SAS Institute, Cary, North Carolina).

RESULTS

Baseline Characteristics

The early group (n = 9941) and the late group (n = 10,313) were well matched with respect to demographic and preoperative comorbidities, with the exception of the following differences: in the early group there were fewer white patients and more patients presenting to the operating room in sepsis, whereas in the late group there were more patients with esophageal varices, a lower mean albumin, and more dialysis patients (Table 2). Of note, there were a similar number of emergency cases in the 2 study groups and the mean operation work RVU (work relative value unit is the metric used in the NSQIP to quantify the complexity of an operation independent of patient preoperative risk factors) was similar between the 2 groups of study.[21] The NSQIP database records an extensive quantity of data regarding laboratory values, and there were no differences ($P > 0.05$) between the 2 groups with respect to bilirubin, creatinine,

TABLE 2. Perioperative Characteristics of Patients in July Compared to the Remainder of the Year

Perioperative Variable	Early Year (N = 9941)	Late Year (N =10,313)	*P*
Demographics			
Male (%)	41.5	41.9	0.541
Race: white (%)	80.8	83.3	<0.001
Mean age (±SD)	54.0 (±17.0)	54.3 (±16.8)	0.145
Operative			
Emergency case (%)	12.9	12.2	0.157
RBCs transfused (±SD)	0.30 (±1.69)	0.27 (±1.51)	0.157
Mean case RVU (±SD)*	15.4 (±9.2)	15.5 (±9.7)	0.336
General			
Albumin (g/dL) (±SD)	3.83 ± 0.7	3.77 ± 0.7	0.001
ASA class ≥3 (%)	37.4	39.1	0.084
Smoker (%)	19.6	19.1	0.424
DNR status (%)	0.38	0.41	0.777
Dependent functional status (%)	1.32	1.57	0.233
Cardiac			
Previous MI (%)	0.93	1.08	0.281
Previous cardiac surgery (%)	6.58	6.36	0.528
Congestive heart failure (%)	1.31	1.29	0.910
Angina (%)	1.67	1.63	0.820
Hypertension (%)	40.1	39.9	0.811
Neurologic			
Previous stroke (%)	2.67	2.49	0.435
History of TIA (%)	2.34	2.18	0.438
Coma (%)	0.18	0.09	0.067
Hemiplegia (%)	1.16	1.16	0.964
Hepatic			
Ascites (%)	1.28	1.34	0.704
Esophageal varices (%)	0.30	0.48	0.048
Nutritional/immune/other			
Diabetes mellitus (%)	12.8	13.0	0.646
Disseminated cancer (%)	3.15	3.38	0.346
Open wound (%)	5.87	5.41	0.152
Steroid use (%)	4.94	4.64	0.326
Weight loss >10% (%)	4.03	4.03	1.0
Bleeding disorder (%)	3.74	3.43	0.236
Sepsis (%)	3.09	2.44	0.005
Pulmonary			
History of COPD (%)	4.05	3.96	0.723
Ventilator >48 h (%)	1.53	1.28	0.132
Dyspnea (%)	13.24	13.98	0.123
Renal			
Acute renal failure (%)	0.61	0.65	0.746
Dialysis (%)	2.30	2.86	0.013
Vascular			
Peripheral vascular disease (%)	4.76	4.92	0.600
Rest pain (%)	3.32	3.57	0.332

*Mean work RVU (Relative Value Units) is the metric for technical effort, stress, mental challenge, and complexity of an operation independent of the preoperative risk factors used by the NSQIP.

hematocrit, platelet count, white blood cell count, sodium, or INR.

Morbidity

The unadjusted 30 day morbidity rate in the early group was 14.3% compared with 13.1% in the late group (P = 0.008). The mean number of complications was 0.24 (±0.73) in the early group compared with 0.22 (±0.70) during the late group (P = 0.016). There were significantly more (unadjusted analysis) postoperative myocardial infarctions (0.42% vs. 0.25%, P = 0.036) and urinary tract infections (3.0% vs. 2.3%, P = 0.003) in the early group. There were no significant differences (unadjusted analysis) in the rate of deep venous thrombosis, pulmonary embolus, wound dehiscence, wound infections (superficial and deep space), renal failure, or stroke.

After multivariable logistic regression to control for potential confounding variables, there was 18% higher risk of postoperative morbidity in the early versus the late group (P = 0.0005). There were 25 variables significantly associated with postoperative morbidity, including ASA, work RVU, albumin, and emergency status of the case, among others (Table 3). Variables which were not included in the model were: coma, pneumonia, history of stroke, diabetes, DNR status, current alcohol abuse, history of congestive heart failure, history of COPD, history of myocardial infarction, on a hypertension medication, platelet count, and sodium level. The discrimination of the model, as measured by a c-index of 0.794, was good.

The expected morbidity rates (derived form a model which excluded time of year as a covariate) showed no seasonal trends in predicted risk of morbidity (based on patient risk factors). To analyze whether the higher morbidity rate during the early group was associated with a trend throughout the year, we analyzed the monthly unadjusted morbidity rates from July to June with a linear regression (Fig. 1). There was a trend toward reduced morbidity as the academic year progressed, although the slope of the line was not significantly different from zero (P = 0.24).

Mortality

The unadjusted 30 day mortality rate in the early group was 2.2% compared the 1.7% in the late group (P = 0.009). After multivariable logistic regression to control for potential confounding variables, there was a 41% higher risk for mortality in the early group compared with the late group. There were 21 other variables significantly associated with higher rates of postoperative mortality including: the American Society of Anesthesiologists physical status classification, albumin, emergency status of the case, and age, among others (Table 4). Other variables which were not in the model included ascites, pneumonia, diabetes, renal failure, current alcohol abuse, history of congestive heart failure, smoking, preoperative steroid use, transfusion, on ventilator preoperatively, wound infection, BMI, and, preoperative hematocrit, and preoperative sodium. The discrimination of the model was excellent, c = 0.938.

The expected mortality rates (derived form a model which excluded time of year as a covariate) showed no

TABLE 3. Multivariable Logistic Regression Model for Risk Factors Associated With Postoperative Morbidity

Step Number	Morbidity Risk Factor	OR	95% CI	*P*	C-Index
1	ASA class 4 or 5 vs. 1 or 2	2.03	1.67–2.47	<0.0001	0.690
1	ASA class 4 or 5 vs. 3	1.32	1.11–1.57	0.0015	0.690
2	Work RVU	1.05	1.05–1.06	<0.0001	0.751
3	Preoperative albumin	0.66	0.60–0.72	<0.0001	0.773
4	Emergency case	1.82	1.59–2.08	<0.0001	0.779
5	Wound infection	1.67	1.42–1.97	<0.0001	0.782
6	Preoperative ventilator	2.07	1.49–2.87	<0.0001	0.783
7	Age	1.01	1.01–1.02	<0.0001	0.784
8	BMI	1.02	1.02–1.03	<0.0001	0.786
9	Smoker	1.33	1.18–1.48	<0.0001	0.787
10	WBC >11,000 cells/mm^3	1.25	1.10–1.41	0.0006	0.788
11	Bleeding disorder	1.28	1.05–1.56	0.0150	0.788
12	Early (vs. late) group	1.18	1.07–1.29	0.0005	0.789
13	Hematocrit <38	1.24	1.11–1.38	0.0002	0.790
14	Partially dependent functional status	1.43	1.20–1.70	<0.0001	0.790
15	Cancer diagnosis	1.36	1.11–1.67	0.0033	0.791
16	Hematocrit >45	1.27	1.06–1.51	0.0090	0.791
17	Dyspnea	1.17	1.04–1.32	0.0104	0.791
18	Transfusion <48 h preoperatively	1.73	1.15–2.58	0.0082	0.792
19	Bilirubin >1.0	1.15	1.02–1.30	0.0232	0.792
20	Ascites	1.47	1.08–1.98	0.0133	0.793
21	Preoperative steroids	1.23	1.03–1.47	0.0240	0.793
22	Dialysis	0.67	0.51–0.87	0.0031	0.793
23	Renal failure	1.60	1.04–2.47	0.0320	0.794
24	Weight loss	1.25	1.03–1.51	0.0229	0.794
25	Creatinine >1.2	1.14	1.00–1.29	0.0441	0.794

Other variables that were not selected in stepwise model building included: coma, pneumonia, history of stroke, diabetes, DNR status, current alcohol abuse, history of congestive heart failure, history of COPD, history of myocardial infarction, platelet count, and sodium level.

*Mean work RVU (Relative Value Units) is the metric for technical effort, stress, mental challenge, and complexity of an operation independent of the preoperative risk factors used by the NSQIP.

significant seasonal trends in predicted risk of mortality (based on patient risk factors). To analyze whether the higher mortality rate during the early group was associated with a trend throughout the year, we analyzed the monthly unadjusted mortality rates from July to June with a linear regression (Fig. 2). There is a significant trend toward lower mortality as the academic year progresses ($P = 0.028$).

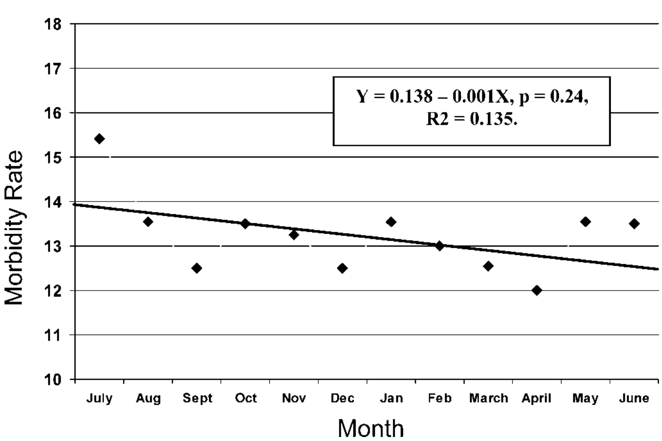

FIGURE 1. Regression line for morbidity over the academic year.

Operating Room Efficiency

We also compared the early group with the late group with respect to operating room efficiency. Fewer cases were done as an outpatient in the early group (Table 5). Cases took 6.3% (8.4 minutes) ($P < 0.001$) longer (period from incision to placement of the dressing) in the early group. Similarly, the time from patient entry into the room to the surgical incision was 5.1% (1.8 minutes) ($P < 0.001$) longer. Finally, the time from the end of the operation to exit from the room was 37.5% (7.4 minutes) ($P < 0.001$) longer. We found an additional 17.4 minutes of OR time per case in the early group.

DISCUSSION

We have demonstrated a significant seasonal variation in surgical morbidity and mortality among centers participating in the ACS-NSQIP. Our results document a progressive reduction in risk-adjusted mortality over the course of the academic year. Although there are other possible explanations for these findings, the cyclic influx of inexperienced

TABLE 4. Multi-variable Logistic Regression Model for Risk Factors Associated With Postoperative Mortality

Step Number	Mortality Risk Factor	OR	95% CI	*P*	C-Index
1	ASA class 4 or 5 vs. 1 or 2	7.30	4.37–12.20	<0.0001	0.862
1	ASA class 4 or 5 vs. 3	2.21	1.64–2.96	<0.0001	0.862
2	Albumin	0.57	0.47–0.68	<0.0001	0.907
3	Emergency case	2.98	2.26–3.92	<0.0001	0.912
4	Age	1.04	1.03–1.05	<0.0001	0.920
5	Platelets <150,000	2.03	1.51–2.72	<0.0001	0.923
6	Totally dependent functional status	2.46	1.67–3.64	<0.0001	0.925
7	DNR status	4.04	2.15–7.58	<0.0001	0.925
8	BUN >40	1.54	1.07–2.20	0.0190	0.926
9	Cancer	2.17	1.41–3.33	0.0004	0.930
10	Work RVU	1.02	1.01–1.03	<0.0001	0.932
11	Dyspnea	1.50	1.15–1.95	0.0028	0.933
12	Coma	5.93	1.86–8.88	0.0026	0.933
13	Weight loss	1.76	1.20–2.59	0.0041	0.935
14	Alkaline phosphatase >125 IU/L	1.51	1.16–1.97	0.0025	0.936
15	Early (vs. late) group	1.41	1.11–1.80	0.0047	0.936
16	Previous MI	2.07	1.30–3.28	0.0020	0.937
17	Creatinine >1.2	1.56	1.18–2.08	0.0020	0.937
18	Hypertension	0.73	0.56–0.95	0.0173	0.938
19	COPD	1.64	1.16–2.32	0.0048	0.938
20	WBC >11,000 cells/mm^3	1.36	1.04–1.79	0.0271	0.938
21	Bleeding disorder	1.43	1.03–1.98	0.0312	0.938
22	Vascular vs. general surgical specialty	0.75	0.56–1.00	0.0484	0.938

Other variables that were not selected in stepwise model building included: ascites, pneumonia, diabetes, renal failure, current alcohol abuse, history of heart failure, smoking, preop steroid use, transfusion, on ventilator pre-operatively, wound infection, BMI, and, pre-op hematocrit, and pre-op Sodium.

*Mean work RVU (Relative Value Units) is the metric for technical effort, stress, mental challenge, and complexity of an operation independent of the preoperative risk factors used by the NSQIP.

trainees each July and the lack of attending the oversight possibly related to vacation schedules, at least in part, underlies observed results. These findings have face validity for anyone who has worked in a busy academic environment.

Our findings are meaningful in that they were generated using the ACS-NSQIP infrastructure, the first standardized risk-adjusted and prospectively collected quality system available for the measurement of surgical outcomes. This system adjusts for the complexity of the operative procedure, the comorbidities of the patients at the time of surgery, and circumstances under which the surgery is performed (emergent vs. nonemergent). That we saw significant differences in month-to-month mortality under these circumstances suggest that there are important factors in academic medical centers that influence mortality, which have not been previously described or accounted for.

Seasonal variation in all cause mortality, usually increased in winter months, has been well documented.[22,23] In the ambulatory (nonsurgical) population, over 75% of the increased mortality is accounted for by deaths from myocar-

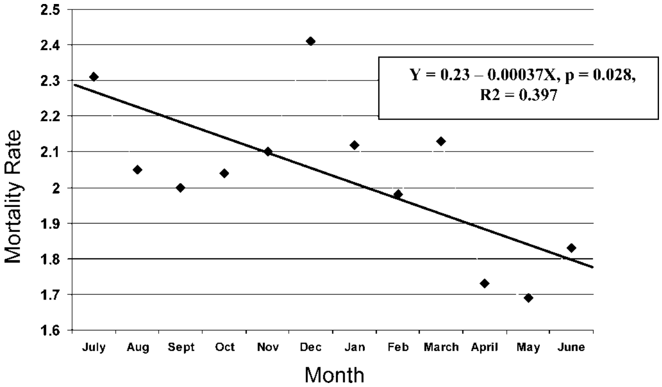

FIGURE 2. Regression line for mortality over the academic year.

TABLE 5. Characteristics of Operations Done in the Early Period of the Academic Year Compared to the End of the Academic Year

	Early Group (n = 9941)	Late Group (n = 10,313)	*P*
Outpatient case (%)	30.3	33.3	<0.0001
Mean OR time (h)	2.36 ± 1.78	2.22 ± 1.72	<0.0001
Mean time patient into room to incision (h)	0.62 ± 0.37	0.59 ± 0.39	<0.0001
End OR to out of room (h)	0.44 ± 0.28	0.32 ± 0.21	<0.0001
Monitored anesthesia (%)	4.2	5.3	0.013
General anesthesia (%)	92.6	91.6	0.013

dial infarction, cerebrovascular disease, and pneumonia. Possible explanations for these findings point to seasonal variation in blood pressure, lipid levels, changes in the activity of the coagulation cascade, and/or to seasonal variation in the prevalence of viral or bacterial pneumonia.[24–27] An underlying theme in these observations is that even small differences in ambient temperature produce activation of the sympathetic nervous system, and a resulting cascade of adverse physiological effects. Such underlying physiologic factors in the general population could theoretically influence mortality after surgical procedures, and these factors are not accounted for in the ACS-NSQIP system. However, our findings demonstrate a dramatic worsening of surgical mortality in July, a summer month, which seems in contrast to the previously reported literature.

A more likely explanation for our findings involves the near universal cycle of changes that occur in academic centers beginning in July. These changes include the influx of new trainees, both medical and nursing, and possibly accentuated by vacation schedules of senior staff during the summer months, resulting in an adequate oversight of the new trainees. However, this study cannot point specifically to the inexperience of trainees in July in explaining the results because almost all of the cohort we studied were academic centers with primary medical school affiliations and anesthesia, general surgery, and vascular surgery training programs. The obvious control group, hospitals without training programs, must be evaluated to determine if a seasonal variation in surgical mortality and morbidity also exists in this setting.

Although the overall slope of the risk-adjusted mortality declined significantly over the course of the academic year, the operative mortality of the month of December was as high as seen in July. Because the month of December is not associated with an influx of new trainees, this spike does not fit the hypothesis that trainee inexperience is solely responsible for our observations. It is, however, consistent with the hypothesis that when complex hospital systems are disrupted, as typically happens in the holiday month of December, operative results suffer.

If the breakdown of complex systems underlies the "July effect," one way to examine this possibility more closely is to evaluate the function of individual process measures that comprise the system in general and subsequently determine whether trainee inexperience results in poor compliance with evidence-based processes. A limitation of this approach, however, and is that the link between process and outcome in medicine, in general, and in surgery specifically, is not well understood.[28] This limitation notwithstanding, we did examine 3 important time intervals in the operative process and compared these time intervals in the "early" versus "late" time periods. Time from operating room entry to skin incision was longer, for example, in "early" versus "late" time periods, possibly reflecting more difficulty for the inexperienced trainees in successfully inserting an endotracheal tube. The duration of surgery was significantly longer in the "early" versus "late" time periods. The time from completion of the procedure to exit from the operating room was 35.5% higher in the "early" versus "late" time periods, a finding which could reflect anesthesia inexperience with the postoperative extubation process. Even thought these modest time differences may not be clinically significant, they to intimate differences in process of care and operating room efficiency early in the academic year.

Although trainee inexperience could account for some of the results we have demonstrated, there are other studies involving cyclic variations in outcomes which implicate different process failures. In surgery, for example, an analysis of the VA-NSQIP, using 112,000 patients, noted increase operative mortality when elective surgical cases were performed on Friday, as opposed to Monday through Wednesday (Zare SM, Itani K, Schifftner TL, et al, unpublished data, 2007). These results do not implicate trainee inexperience. Rather, the results suggest that other factors in hospitals, possibly altered nurse staffing ratios, may affect operative mortality. Nurse staffing ratios have previously been shown to be related to incidence of adverse outcomes in the hospital environment and this could have an influence on our results, with more nursing vacations and associated staffing changes in the July and August.[29–31]

Trainee inexperience may be at the root of our findings, but other possibilities must be examined before firm conclusions can be made. One limitation of our study for example is that case mix is not entirely controlled for in the ACS-NSQIP methodology. Thus, if one accepts that many patients would not elect to have procedures done in the summer, vacation months, the hospitals in question may have experienced a preponderance of urgent, or "semi-elective" cases in July and August, whereas more straightforward, purely elective cases, were performed in the May-June time period. We did control for the "emergent" classification in each time period, but this categorization is very specific for cases done within 12 hours of admission from an emergency department, and it would not capture "semi-elective" and urgent cases. Similarly, many surgeons take vacations in the summer months and this may skew the case mix towards more urgent cases being performed in July and August. Likewise, the RVU measure, which we used to control for differences in case complexity, would not account for the urgency of a particular procedure, complex or not. It seems likely that urgently performed procedures would experience a higher mortality rate 30 days postoperatively, and this skewing of case mix would thus be another explanation for our findings. Ideally, we could tightly control the case mix to operations with relatively time-insensitive indications such as colectomy for colon cancer. Unfortunately, at this time, the sample size in the ACS-NSQIP is insufficient for this type of analysis. Another limitation of this study involves the generalizability of the results to other teaching institutions. The group of hospitals that we have studied represents a relatively uniform cohort of teaching hospitals, namely large academic tertiary care centers. Our study does not address seasonal variation in outcomes in other types of teaching institutions, which may have systems of patient care which rely less on physician trainees.

Our findings are in contrast to previous studies which have suggested that a "July effect" does not exist, including in the Intensive Care Unit environment, or in the care of severely injured patients.[7,13] In a series of 21,679 hospital discharges, it was noted that increasing house officer experience was not associated with hospital mortality, but was associated with a significant decrease in length of stay and total hospital costs.[32] Our study differs from these investigations in that we focused on operative surgical care exclusively, which is more dependent on technical expertise and precise communications between disciplines than other specialties.

Although we studied only academic medical centers, the observations about the "July effect," and postulating that trainee inexperience plays a role in the findings, leads to a broader discussion of quality of care in "teaching" versus "nonteaching" hospitals. Khuri et al addressed this issue in the Veteran's Administration hospitals, and observed that mortality was not different in "teaching" compared with "nonteaching" VA hospitals.[10] In contrast, there were higher complication rates in teaching hospitals, wound infections being the most frequently encountered complication. These observations suggest, but certainly do not prove, that the vulnerabilities of more complex systems of care in an academic setting, which involve more episodes of information transfer, and more frequent changes in the caregivers themselves, could explain the observed results. The question of quality of care in "teaching hospitals" will be more authoritatively answered as the ACS-NSQIP continues to grow. Currently, 150 hospitals participate nationwide, and over half are community hospitals. Because all use the same quality reporting methodology, definitions, and risk adjustment, it should be possible to isolate the single variable of resident involvement to assess the effect of a training program on surgical outcomes in general, and seasonal variation in particular. Fortunately a standardized platform for providing risk-adjusted results exists in the form of the ACS-NSQIP. Using this platform informed questions can be asked and answered.

REFERENCES

1. Birkmeyer JD, Sun Y, Goldfaden A, et al. Volume and process of care in high-risk cancer surgery. *Cancer*. 2006;106:2476–2481.
2. Birkmeyer JD, Stukel TA, Siewers AE, et al. Surgeon volume and operative mortality in the United States. *N Engl J Med*. 2003;349:2117–2127.
3. Birkmeyer JD, Siewers AE, Finlayson EV, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med*. 2002;346:1128–1137.
4. Helmreich RL, Merritt AC, Wilhlem JA. The evolution of crew resource management in training in commercial aviation. *Int J Aviat Psychol*. 1999;9:19–32.
5. Salas E, Burke CS, Bowers CA, et al. Team training in the skies: does crew resource management (CRM) training work? *Hum Factors*. 2001;43:641–674.
6. Morey JC, Simon R, Jay GD, et al. Error reduction and performance improvement in the emergency department through formal teamwork training: evaluation results of the MetTeams project. *Health Serv Res*. 2002;37:1553–1581.
7. Claridge JA, Schulman AM, Sawyer RG, et al. The "July phenomenon" and the care of the severely injured patient: fact or fiction? *Surgery*. 2001;130:346–353.
8. Eappen S, Flanagan H, Bhattacharyya N. Introduction of anesthesia resident trainees to the operating room does not lead to changes in anesthesia-controlled times for efficiency measures. *Anesthesiology*. 2004;101:1210–1214.
9. Fletcher KE, Davis SQ, Underwood W, et al. Systematic review: effects of resident work hours on patient safety. *Ann Intern Med*. 2004;141:851–857.
10. Khuri SF, Najjar SF, Daley J, et al. Comparison of surgical outcomes between teaching and nonteaching hospitals in the Department of Veterans Affairs. *Ann Surg*. 2001;234:370–382; discussion 382–383.
11. Rich EC, Hillson SD, Dowd B, et al. Specialty differences in the 'July Phenomenon' for Twin Cities teaching hospitals. *Med Care*. 1993;31:73–83.
12. Rogers SO Jr, Gawande AA, Kwaan M, et al. Analysis of surgical errors in closed malpractice claims at 4 liability insurers. *Surgery*. 2006;140:25–33.
13. Barry WA, Rosenthal GE. Is there a July phenomenon? The effect of July admission on intensive care mortality and length of stay in teaching hospitals. *J Gen Intern Med*. 2003;18:639–645.
14. Khuri SF, Daley J, Henderson W, et al. Risk adjustment of the postoperative mortality rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. *J Am Coll Surg*. 1997;185:315–327.
15. Khuri SF, Daley J, Henderson W, et al. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. *Ann Surg*. 1998;228:491–507.
16. Khuri SF. The NSQIP: a new frontier in surgery. *Surgery*. 2005;138:837–843.
17. Dimick JB, Chen SL, Taheri PA, et al. Hospital costs associated with surgical complications: a report from the private-sector National Surgical Quality Improvement Program. *J Am Coll Surg*. 2004;199:531–537.
18. Fink AS, Campbell DA Jr, Mentzer RM Jr, et al. The National Surgical Quality Improvement Program in non-veterans administration hospitals: initial demonstration of feasibility. *Ann Surg*. 2002;236:344–353; discussion 353–354.
19. Hua HT, Cambria RP, Chuang SK, et al. Early outcomes of endovascular versus open abdominal aortic aneurysm repair in the National Surgical Quality Improvement Program-Private Sector (NSQIP-PS). *J Vasc Surg*. 2005;41:382–389.
20. American College of Surgeons-National Surgical Quality Improvement Program website. Data collection form. Available at: www.acsnsqip. Accessed July 19, 2006.
21. Aust JB, Henderson W, Khuri S, et al. The impact of operative complexity on patient risk factors. *Ann Surg*. 2005;241:1024–1027; discussion 1027–1028.
22. van Rossum CT, Shipley MJ, Hemingway H, et al. Seasonal variation in cause-specific mortality: are there high-risk groups? 25-year follow-up of civil servants from the first Whitehall study. *Int J Epidemiol*. 2001;30:1109–1116.
23. Gemmell I, McLoone P, Boddy FA, et al. Seasonal variation in mortality in Scotland. *Int J Epidemiol*. 2000;29:274–279.
24. Brennan PJ, Greenberg G, Miall WE, et al. Seasonal variation in arterial blood pressure. *Br Med J (Clin Res Ed)*. 1982;285:919–923.
25. Stout RW, Crawford V. Seasonal variations in fibrinogen concentrations among elderly people. *Lancet*. 1991;338:9–13.
26. Stout RW, Crawford VL. Seasonal variations of plasma fibrinogen in elderly people. *Lancet*. 1994;343:975.
27. Eastridge BJ, Hamilton EC, O'Keefe GE, et al. Effect of sleep deprivation on the performance of simulated laparoscopic surgical skill. *Am J Surg*. 2003;186:169–174.
28. Bradley EH, Herrin J, Elbel B, et al. Hospital quality for acute myocardial infarction: correlation among process measures and relationship with short-term mortality. *JAMA*. 2006;296:72–78.
29. Aiken LH, Clarke SP, Sloane DM, et al. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA*. 2002;288:1987–1993.
30. Aiken LH, Clarke SP, Cheung RB, et al. Educational levels of hospital nurses and surgical patient mortality. *JAMA*. 2003;290:1617–1623.
31. Dimick JB, Swoboda SM, Pronovost PJ, et al. Effect of nurse-to-patient ratio in the intensive care unit on pulmonary complications and resource use after hepatectomy. *Am J Crit Care*. 2001;10:376–382.
32. Rich EC, Gifford G, Luxenberg M, et al. The relationship of house staff experience to the cost and quality of inpatient care. *JAMA*. 1990;263:953–957.

Discussions

DR. THOMAS R. RUSSELL (CHICAGO, ILLINOIS): I think the NSQIP has great opportunities, and I think this study points out some of the uses of the data.

However, as you point out in the paper, there are limitations with the information that you came up with, and clearly you describe in the paper a phenomenon that you identified. I think we have to be very careful with data like this, because this points out a variability of care that deals with the timing of surgery and the so-called July phenomenon.

Thinking of all the papers that were presented here the last day and a half, I can think of no paper that would be of more interest to the public or to the press than this paper. Perhaps I am over sensitized to this since I am often called by the press on various issues. So this once again emphasizes the importance of being so careful with data and the way it is presented. I could just see the press getting hold of something like this and suddenly the public is presented with "It is very dangerous to be operated on in the month of July," and it would have huge implications for health care, once again feeding the idea of the variability of care. So I really commend you for doing this, and also caution you in going forward with this preliminary information, and the need to have it truly validated.

I have 1 question for you. You mentioned that in this study you had a small group of community hospitals. Do you have enough data to give us any information whether this phenomenon is at all seen in other hospitals where perhaps many services are not performed so heavily by house staff?

I enjoyed the paper. And once again I emphasize the importance of these large databases to come to some conclusion about health care. And I think this paper nicely pointed that out.

DR. MICHAEL J. ENGLESBE (ANN ARBOR, MICHIGAN): In short, the answer is no. We had about 120,000 cases in our data set and they contributed less than 2000 of those cases. That was just because of the nature of the study. It was a 3-year study and they were only enrolled in the final year of the study.

To add to that, though, there are lots of important studies we need to do based on this preliminary study. And I think we need to look at other data sets, which we are in the process of doing, such as Medicare where we can control better for case mix. But, I think the study that I and Dr. Campbell are excited to do is to use the newer NSQIP data. For those of you not familiar with the NSQIP, it started in the VA, when it had only 14 academic hospitals. It now has over 150 hospitals in the United States, over half of which are community-based hospitals. That data is being collected as we speak, and hopefully within the next year there will be adequate data with which we can repeat the analysis and improve on the analysis with a control group.

DR. KEITH D. LILLEMOE (BALTIMORE, MARYLAND): As anyone who sits through morbidity and mortality conferences realizes, many times there is the absence of the attending physician at the time of the patient's disastrous complication. I think we are even aware of a very prominent lawsuit that was recently dismissed related to somebody being away. July and August are peak vacation times for most of us in academic surgery. The peak at Christmas time in December may influence this. How can you separate the supervision, the hand-off aspects, away from just the inexperience of the surgical residents?

DR. MICHAEL J. ENGLESBE (ANN ARBOR, MICHIGAN): I think that is a very important question, and it is the exact question that John Birkmeyer asked us when we initially showed him this data. And we cannot. In our current data set, we cannot separate the two. The fact that we see a big bump in December is certainly interesting, we did not analyze it thoroughly, but it certainly could be related to less oversight, fewer attending physicians in the hospital, so to speak.

I think the way around that is to tightly control case mix as much as possible. For example, look at operations such as cavage or carotid endarterectomy where we could see a large number of operations, and I think theoretically there would be less variation. I think it is quite simply, you have a cardiac surgery Fellow managing the operation or managing the postoperative care, and there may be less variation, as opposed to some of these general surgery operations.

But I do not have a good answer. And I think it is something that we will have to work hard on to try to get around. Because it certainly is likely a very profound confounder, and that is one of the primary reasons we need to be cautious about how we interpret the data.

DR. THOMAS R. GADACZ (AUGUSTA, GEORGIA): I share Dr. Lillemoe's feelings about some of the implications. I think it is very easy to jump to the assumption that you have new house staff and a loss of some seasoned house staff in July, to attribute problems to this turnover.

I think this is a very dangerous jump to make. To me, you seem to want to fit everything into the fact that these are inexperienced people. And if you actually look at your data, I don't really think that there is a basis for your claim, because there should not be an increase in mortality in December.

I think the way I would interpret this data is not so much as a lack of experience but as a lack of supervision. And this really isn't a problem with process of care; this is really a problem of adequate supervision during all stages. The peak seasons for vacations are actually in summer and in December. So to me, I would be concerned about adequate supervision during these times when you observed an increase in mortality.

Dr. Michael J. Englesbe (Ann Arbor, Michigan): I agree. And I want to make it clear that we are not standing here telling you that the cause of this phenomenon is physician inexperience. We are trying to be cautious about doing that. We have not proven anything. We have shown a phenomenon.

Dr. Jean C. Emond (New York, New York): When I was a Fellow in France, I was amazed by the month of August over there. Most of the doctors take their vacation in August. They shut the hospitals nearly completely, they shut ICUs. So in the end, the only patients who show up for care are critically ill.

You alluded, very wisely, to the limitations in your study suggesting, and we tend to focus on trainees, on our own supervision. But I think the patients change quite a bit. It was interesting in your complication rates that the only significant complication rates were UCIs and MRIs, which are not what I usually would associate with less experienced docs.

I wanted to ask you about another issue, though, which is, in New York we face the work hours very early and we never fully recover. Many of our services no longer get residents at all. And to the extent that we have transitioned the physician extenders, it tends to occur on the higher CMI-type services, transplant, heart and so forth. What would the impact be, in your thoughts, of transitioning to physician extenders with year-round consistency in knowledge and care in the setting of some of your observations?

Dr. Michael J. Englesbe (Ann Arbor, Michigan): If we decide as a surgical group that this is physician inexperience or physician vacations –it doesn't matter what it is, there are mortality differences–I think things like physician extenders really may be an excellent way to bridge transitions in care with problems with communication. Experienced physician extenders, I think, are great assets to inexperienced physicians, or are potentially a great transition in times when staffing is low such as during vacation times. So I think physician extenders potentially could play an important role in this.

Dr. H. Gill Cryer (Los Angeles, California): This was a very provocative study and very interesting data. I have a couple of questions for you.

First of all, when the NSQIP group tried to analyze the reasons for variability in outcome between what they called high-performing centers and centers for concern in their initial analysis, almost all of the reasons they found when they went to the individual institutions were structural or system problems–lack of a good performance improvement program, lack of resources or scheduling or guidelines or protocols–and very few, if any, of the reasons for variability were in physician decisions, judgments or care.

Along those lines, I think that regarding the vacation periods, it is not just the physicians who go on vacation but it is everybody in the hospital, the nurses, the x-ray techs, the radiologists, and indeed the patients even want to go on vacation. So, only the high-risk patients that must have an operation have them during that time. And I would ask you to consider that in your analysis.

Secondly, there is 1 group of patients that is fairly highly regulated along the lines of these process and system issues, and those are trauma patients and trauma centers. And several of the institutions you examined were Level I trauma centers. Did you notice any difference? And if you look at specifically trauma patient care, was there the same variation in those where structures and systems are highly regulated compared to other patients?

Dr. Michael J. Englesbe (Ann Arbor, Michigan): That is a good point. There were not many trauma patients in the study in general, mostly general vascular surgery cases. Twelve percent of the operations were emergent in both groups. I think maybe some of the robust trauma registries could provide a great opportunity to conduct a similar analysis to see if there are seasonal variations.

When we have done this in Medicare, though, for example looking at hip surgery for hip fracture, we noticed that there is a true seasonal variation in mortality with that probably related to the severity of the hip fracture. And that may also be a problem with any trauma patient analysis. I think there is generally more trauma, maybe even more severe trauma, earlier, at different times of year. So it is difficult.

Dr. Andrew L. Warshaw (Boston, Massachusetts): As a matter of interest, of the hospitals in your study, 5 were from our system, that is, the 2 Partners academic centers, MGH and Brigham, and the 3 community hospitals. Just as a point of information, during the time of that study and up to the present time we have carefully compared the 5 institutions within the group, and have found no differences between the community hospitals and the 2 academic centers using NSQIP data, even though the presence of residents covering cases at the community hospitals in our system is much less. The "resident factor," if there is one, is not apparent.

I have 2 questions for you: First, morbidity was not significantly different for July compared with the later time periods while mortality was. That discrepancy is counterintuitive. One would expect more in the way of complications in a population that was destined to die. How do you explain the seeming paradox?

My second question is about the increased operating time, 17 minutes longer in July. While that may be statistically significant, is it clinically significant?

DR. MICHAEL J. ENGLESBE (ANN ARBOR, MICHIGAN): To answer your second question first, I think that 17 minutes longer time in the operating room is clinically significant for 2 reasons. One, there are lots of data showing the longer the operation the higher the risk of wound infection, DVT, things like that. But I think even more importantly, it may intimate some difficulty with the operation. The operation takes longer when you are operating with somebody with less experience, more intraoperative, maybe not true complication but inconveniences could be happening. Certainly I think it does happen, in my experience, early in the academic year. So I think differences in the operation are clinically significant.

I think the really exciting thing is this difference in depth of anesthesia. And we are really looking hard at this. Somebody that has too much anesthesia on-board can experience episodes of hypotension. And we now have–actually 1 of the hospitals at Mass General will bring them, in addition to the University of Michigan–these interoperative modern systems where we are actually in the process of doing the analysis, looking at whether the depth of anesthesia varies on a seasonal basis. Presumably that could be related to physician inexperience.

DR. EDWARD M. COPELAND, III (GAINESVILLE, FLORIDA): What I am going to say may already have been said in a different format, but this is the first time that I have heard a paper presented with the American of College Surgeons NSQIP data. And I want to speak in my capacity as president of the American College of Surgeons and I want to acknowledge the hard work that Scott Jones and Karen Richards did virtually single-handedly– or maybe double-handily is the way to put it–to bring the NSQIP program together for all of us to have the data to work with. This is preliminary, but let me tell you, the people who obtained the data for you to look at worked as hard as I have seen any 2 people work in my career. So, Scott, if you are in the room, thank you so much for having done that. Karen is here. I do not think she is probably in the room, but I would like to acknowledge her efforts as well. Clifford Ko will take this over and I am sure will do an equally good job. But since this is recorded for posterity, I would like to have that piece of information in the record.

DR. DARRELL A. CAMPBELL, JR. (ANN ARBOR, MICHIGAN): While many of the previous discussants have focused, I think appropriately, on the idea that it is systems of care that we need to be looking at, it is possible that residents and resident density per bed at some point influenced those systems of care.

But I just want to point out that there is a very important paper that Shukri Khuri, another pioneer in NSQIP, has recently published. It showed that the odds ratio for mortality in patients that were operated upon in the VA system on Fridays as opposed to Monday, Tuesday and Wednesday, was about 1.3, very similar to our study. So this hardly implicates resident experience in that kind of phenomenon. That is a system phenomenon. It may have something to do with nurses per bed or something like that on the weekends, or supervision on the weekends, but it is a system issue that is very independent from residents, possibly. And I think it just shows that we need to dive deeper into this interesting finding that we have seen, but focus on systems.